CSC 423 – A3

Devroop Banerjee
V00837868

1. **Exercise 4.1**
Alice and Bob play checkers often. Alice is a better player, so the probability that she wins any given game is 0.6, independent of all other games. They decide to play a tournament of n games. Bound the probability that Alice loses the tournament using a Chernoff bound.

Ex 4.1: V00837868

Alice looses the game if she wins less than half the total # of games played, with $Pr(x < \frac{n-1}{2})$

Use Chernoff bound: $Pr(x \le (1-\delta)\mu) \le e^{-(\mu\delta^2)/2}$

Solve for $\mu = 3n/5$ for the inequality $(1-\delta)\mu \geq \frac{n-1}{2}$

which gives $\delta \leq 1/3$ for $n \geq 5$. $\Rightarrow (1-\delta)(3n/5) \geq \frac{n-1}{2}$

We can use $\delta \leq \alpha$, $\forall \alpha > 1/6$ $\Rightarrow \delta \leq 1 - \frac{5n-1}{6n}$

and sufficiently large n.

$$Pr(x \leq (n-1)/2) \leq Pr(x \leq (1-\delta)\mu) \leq e^{-(\mu\delta^2)/2}$$
$$= e^{-((3n/5)(1/9))/2}$$
$$= e^{-n/20}.$$

2. **Exercise 4.2**
We have a standard six-sided die. Let *X* be the number of times that a 6 occurs over *n* throws of the die. Let *p* be the probability of the event $X \geq n/4$. Compare the best upper bounds on *p* that you can obtain using Markov's inequality, Chebyshev's inequality, and Chernoff bounds.

Ex 4.2:

- Chernoff: $p = Pr(x \geq n/4) = Pr(x \geq (1+\delta)\mu) \leq e^{-(\mu\delta^2)/3}$
Since $E(x) = \mu = n/6$, solve $(1+\delta)\mu = n/4$ to obtain $\delta = 1/2$. We can bound p as $p \leq e^{-((n/6)(1/4))/3} = e^{-n/72}$

- Markov: $P(x \geq t) \leq \left(\frac{E(x)}{t}\right)$ where $E(x) = E(\Sigma x_i) = n/6$.
We can deduce that $p = P(x > n/4) \leq \frac{(n/6)}{(n/4)}$
$= 2/3$

- Chebyshev: $P(x \geq t) \leq P(|x - E(x)| \geq t) \leq \frac{Var(x)}{t^2}$
Since $Var(x) = nq(1-q)$, we get
$$P(x \geq n/4) \leq P(|x - E(x)| \geq n/4) \leq \frac{\left(\frac{5n}{36}\right)}{\left(\frac{n}{4}\right)^2}$$
$$= \frac{20}{9n}$$

3. **Exercise 4.10**
A casino is testing a new class of simple slot machines. Each game, the player puts in $1, and the slot machine is supposed to return either $3 to the player with probability 4/25, $100 with probability 1/200, or nothing with all remaining probability. Each game is supposed to be independent of other games.
The casino has been surprised to find in testing that the machines have lost $10,000 over the first million games. Derive a Chernoff bound for the probability of this event. You may want to use a calculator or program to help you choose appropriate values as you derive your bound.

Ex 4.10:

Let rv x denote net loss to the casino over 1st million games. $x = x_1 + x_2 + \dots + x_{10^6}$.

Let x_i for $i = 1, 2, \dots, 10^6$ be the casino's net loss in the i^{th} game.

$$x_i = \begin{cases} 2 & \text{w.p. } 4/25 \\ 99 & \text{w.p. } 1/200 \\ -1 & \text{w.p. } 167/200 \end{cases} \Rightarrow e^{tx_i} = \begin{cases} e^{2t} & \text{w.p. } 4/25 \\ e^{99t} & \text{w.p. } 1/200 \\ e^{-1t} & \text{w.p. } 167/200 \end{cases}$$

$$\therefore E(e^{tx_i}) = \tfrac{4}{25} e^{2t} + \tfrac{1}{200} e^{99t} + \tfrac{167}{200} e^{-t}$$

$$E(e^{tx}) = E(e^{t(x_1 + x_2 + \dots x_{10^6})})$$

$$= E(e^{tx_1}) \cdot E(e^{tx_2}) \cdot \dots \cdot E(e^{tx_{10^6}}) \quad (x_i\text{'s are independent})$$

$$= \left(\tfrac{4}{25} e^{2t} + \tfrac{1}{200} e^{99t} + \tfrac{167}{200} e^{-t}\right)^{10^6}$$

$$Pr(x > 10^4) = Pr(e^{tx} \geq e^{10^4 t})$$

$$\leq \frac{E(e^{tx})}{e^{10^4 t}} \quad (\text{by Markov's})$$

$$= \left(\tfrac{4}{25} e^{2t} + \tfrac{1}{200} e^{99t} + \tfrac{167}{200} e^{-t}\right)^{10^6} \cdot e^{-10^4 t}$$

This bound holds for any $t > 0$, so we can choose the best value for t. Let's pick a small value like $t = 0.0006$, which gives us a bound of 0.0002, suggesting that the casino has faulty machines.

4. **Exercise 4.12**
Consider a collection *X*1, . . . , *Xn* of *n* independent geometrically distributed random variables with mean 2. Let *X* =_*ni*=1 *Xi* and *δ* > 0.

(a) Derive a bound on Pr(*X* ≥ (1 + *δ*)(2*n*)) by applying the Chernoff bound to a sequence of (1 + *δ*)(2*n*) fair coin tosses.

Ex 4.12:

a) x_i: Sequence of coin flips till 1^{st} head

Σx_i: " " " " " n^{th} "

$x > (1+\delta)2n$: n^{th} head doesn't occur in 1^{st} $(1+\delta)2n$ flips.

$\Pr(x > (1+\delta)2n) = \Pr(Y < n)$

$* \; E[Y] = (1+\delta)n \; *$

$$\begin{aligned} \Pr(x > (1+\delta)2n) &= \Pr(Y < n) \\ &= \Pr\left(Y < \left(1 - \frac{\delta}{1+\delta}\right)(1+\delta)n\right) \\ &\le e^{-\left((1+\delta)n \cdot \frac{\delta^2}{2(1+\delta)^2}\right)} \\ &= e^{\left(\frac{n\delta^2}{2(1+\delta)}\right)} \end{aligned}$$

(b) Directly derive a Chernoff bound on Pr(X ≥ (1 + δ)(2n)) using the moment generating function for geometric random variables.

b) $e^t < 2$,

$$\begin{aligned} \therefore E[e^{tx_i}] &= \tfrac{1}{2}e^t + \tfrac{1}{4}e^{2t} + \tfrac{1}{8}e^{3t} + \dots \\ &= \textstyle\sum_{k=1}^{\infty} \left(\frac{e^t}{2}\right)^k \\ &= \frac{e^t/2}{1-(e^t/2)} \\ &= \frac{e^t}{2-e^t} \end{aligned}$$

Using Markov's inequality,

$$\begin{aligned} \Pr(x > (1+\delta)2n) &= \Pr\left(e^{tx} > e^{(t(1+\delta)2n)}\right) \\ &\le \frac{E(e^{tx})}{e^{(t(1+\delta)2n)}} \\ &= \frac{\prod E(e^{tx_i})}{e^{(t(1+\delta)2n)}} \\ &\le \frac{e^{tn}}{(2-e^t)e^{(t(1+\delta)2n)}} \qquad \left(t = \ln\left(1 + \frac{\delta}{1+\delta}\right)\right) \\ &= \left(\left(1 - \frac{\delta}{1+\delta}\right)\left(1 + \frac{\delta}{1+\delta}\right)^{1+2\delta}\right)^{-n} \\ &\le e^{\left(-\epsilon + (1+\epsilon)\left(\frac{\delta^2}{1+\delta} + \delta\right)\right)-n} \qquad (\text{when } \delta \text{ is small}, \exists \; \epsilon > 0) \\ &\le e^{-(n((1-\epsilon)\delta^2 - \epsilon))} \end{aligned}$$

(c) Which bound is better?

c) When δ is small, st ϵ is close to 0, the bound in (b) is tighter than the one in (a).

5. **Exercise 5.10**

Consider the probability that every bin receives exactly one ball when
n balls are thrown randomly into *n* bins.

(a) Give an upper bound on this probability using the Poisson approximation.

Ex 5.10:

a) Let Y_i be a Poisson r.v. for bin i. Also $m = n$.

$Pr(Y_i = j) = \frac{e^{-n/n}(\frac{n}{n})^j}{j!}$

$Pr(Y_i = 1) = 1/e$

By Applying Theorem 5.7, we can bound the Pr. of X using $Y = \{\sum_{i=1}^{m} Y_i = 1\}$

$Pr(X) \leq e\sqrt{n}\, Pr(Y)$

$= e\sqrt{n}\,(1/e)^n$

(b) Determine the *exact* probability of this event.

b) There are n^n ways to throw n balls into n bins. There's only 1 way (given an ordering) to put a single ball into each bin. There are $n!$ possible orderings. Thus exact prob of n balls in n bins $\Rightarrow Pr(X) = \frac{n!}{n^n}$

(c) Show that these two probabilities differ by a multiplicative factor that equals the probability that a Poisson random variable with parameter *n* takes on the value *n*. Explain why this is implied by Theorem 5.6

6. **Exercise 5.14**

We prove that if Z is a Poisson random variable of mean μ, where
$\mu \geq 1$ is an integer, then $Pr(Z \geq \mu) \geq 1/2$.

(a) Show that $Pr(Z = \mu + h) \geq Pr(Z = \mu - h - 1)$ for $0 \leq h \leq \mu - 1$.

Ex 5.14:

a) $P(Z = \mu + h) \geq P(Z = \mu - h - 1)$ for $0 \leq h \leq \mu - 1$

$$\frac{e^{-\mu}\mu^{(\mu+h)}}{(\mu+h)!} \geq \frac{e^{-\mu}\mu^{(\mu-h-1)}}{(\mu-h-1)!}$$

$$\frac{\mu^{(\mu+h)}}{(\mu+h)!} \geq \frac{\mu^{(\mu-h-1)}}{(\mu-h-1)!}$$

$$\mu^{2h+1} \geq \frac{(\mu+h)!}{(\mu-h-1)!}$$

$$\mu^{2h+1} \geq (\mu-h)(\mu-h+1)\ldots(\mu-1)\mu(\mu+1)\ldots(\mu+h-1)(\mu+h)$$

Multiply pairs of co-effs $(\mu-h)(\mu+h),\ldots,(\mu-1)(\mu+1)$. There are h of these co-effs $(\mu-(h-i))(\mu+(h-i)) = (\mu^2-(h-i)^2)$, $(\mu^2-(h-i)^2) \leq \mu^2$. Inequality holds for $h \geq 0$, since:

$$\mu^{2h+1} \geq (\mu^2-h^2)(\mu^2-(h-1)^2)\ldots(\mu^2-1).\mu$$

$$\mu^{2h} \geq (\mu^2-h^2)(\mu^2-(h-1)^2)\ldots(\mu^2-1).$$

(b) Using part (a), argue that Pr($Z \geq \mu$) ≥ 1/2.

b) Using part (a) we can argue that $P(Z \geq \mu) \geq \frac{1}{2}$.

$$\begin{aligned} P(Z \geq \mu) &= \sum_{h=\mu}^{\infty} P(Z = h) \\ &= \sum_{h=0}^{\infty} P(Z = \mu + h) \\ &\geq \sum_{h=0}^{\mu-1} P(Z = \mu + h) \\ &\geq \sum_{h=0}^{\mu-1} P(Z = \mu - h - 1) \text{ (by part (a))} \\ &= P(Z < \mu) \end{aligned}$$

Since $P(Z < \mu) + P(Z \geq \mu) = 1$ & $P(Z \geq \mu) \geq P(Z < \mu)$ from above, we get $P(Z \geq \mu) \geq \frac{1}{2}$.

7. **<u>Exercise 5.22</u>**

In hashing with *open addressing,* the hash table is implemented as an array and there are no linked lists or chaining. Each entry in the array either contains one hashed item or is empty. The hash function defines, for each key *k*, a *probe sequence* $h(k, 0)$, $h(k, 1)$, . . . of table locations. To insert the key *k*, we first examine the sequence of table locations in the order defined by the key's probe sequence until we find an empty location; then we insert the item at that position. When searching for an item in the hash table, we examine the sequence of table locations in the order defined by the key's probe sequence until either the item is found or we have found an empty location in the sequence.

If an empty location is found, this means the item is not present in the table.

An open-address hash table with 2*n* entries is used to store *n* items. Assume that the table location *h*(*k, j*) is uniform over the 2*n* possible table locations and that all *h*(*k, j*) are independent.

(a) Show that, under these conditions, the probability of an insertion requiring more than *k* probes is at most 2−*k*.

Ex 5.22:

a) Consider i^{th} iteration. Since $i-1$ entries are already filled, Success prob $= \frac{2n-(i-1)}{2n} > \frac{1}{2}$.
Thus, at each step with prob atleast $\frac{1}{2}$, find an empty entry.
Since $h(k,j)$ are iid, $Pr(\text{More than } m \text{ probes}) \leq \prod_{i=1}^{m} \frac{1}{2} = 2^{-m}$.

(b) Show that, for *i* = 1, 2, . . . , *n*, the probability that the *i*th insertion requires more than 2 log *n* probes is at most 1/*n*2.

b) Using part (a), we get $Pr(x_i > m) \leq 2^{-m} \Rightarrow$
$Pr(x_i > 2\log n) \leq 2^{-2\log n} = n^{-2}$

Let the random variable *Xi* denote the number of probes required by the *i*th insertion. You have shown in part (b) that Pr(*Xi* > 2 log *n*) ≤ 1/*n*2. Let the random variable *X* = max1≤*i*≤*n Xi* denote the maximum number of probes required by any of the *n* insertions.

(c) Show that Pr(*X* > 2 log *n*) ≤ 1/*n*.

(d) Show that the expected length of the longest probe sequence is **E**[*X*] = *O*(log *n*).